Object recognition: **Seeing us seeing shapes**

James A. Mazer and Jack L. Gallant

Our understanding of the neural basis of object recognition is based primarily on work with non-human primates. The problem has recently been addressed in humans using functional magnetic resonance imaging; new results indicate that the lateral occipital complex plays an important role in human object recognition.

Address: Department of Psychology, University of California, Berkeley, 3210 Tolman Hall #1650, University of California at Berkeley, Berkeley, California 94720-1650, USA.
E-mail: mazer@socrates.berkeley.edu, gallant@socrates.berkeley.edu

Current Biology 2000, 10:R668–R670

0960-9822/00/$ – see front matter

Object recognition is a seemingly effortless process that requires no conscious thought. For more than 100 years, psychologists and neuroscientists have sought to understand the neural mechanisms underlying our innate ability to perceive and identify objects, using techniques ranging from psychophysical studies of human subjects to single neuron recording from non-human primates. Despite these efforts, the neural substrates of object recognition remain elusive. Insights into the neural circuits that underlie object recognition in humans are now coming from studies using functional magnetic resonance imaging (fMRI). One recent study [1] has provided evidence that one area of the visual cortex, the lateral occipital complex (LOC), plays an important role in human object recognition — the representation of complex shapes.

Object recognition during natural vision is a difficult computational problem. A complex natural scene is actually a collection of surface patches differing in color, luminance, texture and shape. To recognize a single object, the constituent patches must be segmented from the background and grouped together into a coherent whole. The geometrical relationships between these components may also need to be computed to identify the objects' three-dimensional structure. The object must then be identified on the basis of past experience.

In non-human primates, object recognition is mediated by a network of over 30 distinct cortical areas, arranged in a multilayered hierarchy (for review see [2]). Information moves through the system both bottom-up — from the eye to the top of the hierarchy — and top-down. A series of ventrally located cortical areas are thought to be primarily responsible for object recognition. Visual input reaches the ventral stream via a thalamocortical projection from the dorsal lateral geniculate nucleus to the primary visual cortex (area V1). V1 in turn projects to V2, V2 to V4, and V4 to inferotemporal (IT) cortex, which itself probably includes several distinct areas that mediate object vision.

The human and macaque ventral streams appear to be structurally and functionally homologous up to the level of V4 [3]. In humans, fMRI studies suggest that several areas beyond V4 might participate in object recognition. Potential higher visual areas in the human brain include the LOC [4], the fusiform face area [5] and the parahippocampal place area [6]. These candidate areas are clustered together near the temporal-occipital junction, and are all active during object recognition. The exact relationship between these areas and macaque IT has yet to be resolved.

Visual function is typically studied by parametrically varying a single attribute of the visual scene, while keeping all other attributes constant. Correlations between attribute variations and neural activity provide information about areas or neural circuits involved in perception of the attribute of interest. In the case of luminance and color, which are easily quantified, this approach works well. Visual neuroscientists using this strategy have made substantial progress in identifying the specific brain structures and circuits responsible for relatively simple perceptual capacities, such as luminance and color perception and the ability to see simple patterns. Unlike luminance, complex shapes and three-dimensional objects cannot easily be quantified and parameterized. As a result, neuroscientists have been less successful at understanding complex shape processing and object recognition in visual areas beyond V4.

To circumvent the need for a parametric 'shape-space', some researchers have recently focused on simply identifying cortical areas specifically involved in the recognition of complex objects. Malach *et al.* [4] asked observers to view a set of photographs while measuring brain activity with fMRI. Some of the photographs depicted recognizable objects, while others were degraded by noise so as to be unrecognizable. Malach and colleagues identified a specific region of human visual cortex, the LOC, which was activated only by recognizable objects. The LOC is anatomically situated along the posterior portion of the fusiform gyrus (anterior to Brodmann's area 19).

Demonstrating that a specific cortical region responds to visual objects is necessary, but not sufficient, for declaring it to be the site of object recognition. Activation during object viewing could be due to other processes, such as visual attention or arousal. Therefore, additional

experiments are required to demonstrate that LOC activity reflects object recognition.

Theorists have proposed that a robust object recognition system should be relatively insensitive to the precise physical cues that denote an object; an apple should be recognized as an apple whether it is a real apple, a photograph of an apple or a drawing of an apple. Insensitivity to the specific visual cues that comprise an object is known as 'form-cue invariance'. A good object recognition system should also be insensitive to other transformations that do not change an objects' identity, such as translation and scaling; this is called 'perceptual constancy'. Neurons in IT of the macaque appear to satisfy both of these requirements; they are highly selective for particular shapes and exhibit both form-cue invariance and perceptual constancy with respect to changes in stimulus position and size (for review see [7]).

Several laboratories have used fMRI to identify human cortical areas that exhibit form-cue invariance and perceptual constancy. Grill-Spector *et al.* [8] asked observers to passively view objects defined either by luminance cues (silhouette figures) or motion cues (silhouettes with the foreground composed of dots moving in one direction against a background of stationary dots) while using fMRI to measure blood oxygenation level-dependent (BOLD) activation. The LOC was activated regardless of which cues were used to denote shape, confirming form-cue invariance.

Malach *et al.* [4] demonstrated that the LOC shows perceptual constancy with respect to changes in stimulus size. The origin of the observed constancy was unclear, however, because of the limited spatial resolution of conventional BOLD fMRI. The size constancy observed by Malach *et al.* [4] could reflect size invariance in single LOC neurons, as reported for the macaque IT cortex [9]. Alternatively, the data could reflect the average response of a diverse population of neurons, each tuned for a different object–size combination. In this case, no single neuron would be perceptually constant across all sizes.

Grill-Spector *et al.* [10] addressed this issue by examining adaptation of the BOLD response. Previous studies demonstrated that repetitive stimulus presentation reduces both neural activity [11] and the BOLD response [12], and that normal response levels are restored when a novel stimulus is presented. Grill-Spector *et al.* [10] used faces that varied in size, position, viewpoint and illumination. To test size invariance, they measured BOLD adaptation in the LOC while changing image size and holding position, viewpoint and illumination constant (analogous experiments were used to assess the other constancies). They reasoned that, if single LOC neurons were size constant, then they would adapt to repeated presentations of the same stimulus, even if that stimulus varies in size. But if different neurons encode the same objects at different sizes, then changing stimulus size would block adaptation.

Grill-Spector *et al.* [10] found adaptation of the BOLD response in the anterior–ventral portion of the LOC, termed LOa, when a single image was presented repeatedly at different sizes or positions. The caudal–dorsal portion of the LOC, termed LO, while clearly activated by objects, did not exhibit size or position invariant adaptation. Neither LO nor LOa appeared to be viewpoint or illumination invariant. One interpretation of these results is that LO contains distinct populations of shape-selective neurons tuned to different scales and positions, while single neurons in LOa integrate over scale and position, possibly by summing inputs from LO. In general, these results support a model in which perceptual constancy in LOa occurs at the single neuron level, or at least at a level below the spatial resolution of fMRI.

More recently, Kourtzi and Kanwisher [1] examined the specific visual cues used by the LOC to define objects. They measured the BOLD signal in the LOC while observers viewed photographs and line drawings of novel abstract three-dimensional objects. They found that repetitive viewing of photographs, line drawings or a mixture of the two led to equal BOLD adaption in the LOC. This finding is consistent with the results of Grill-Spector *et al.* [10] and suggests that form-cue invariance, like size and position constancy, also occurs at the level of the single neuron or below fMRI resolution.

Kourtzi and Kanwisher [1] also measured BOLD responses to three different stimulus classes: line drawings of three-dimensional objects, line drawings of partially occluded two-dimensional shapes that have depth but no three-dimensional structure, and simple two-dimensional line drawings with neither depth nor three-dimensional structure. The observed LOC activity was similar in all three viewing conditions. Taken together, the results reported by Kourtzi and Kanwisher [1] demonstrate that the LOC is not merely specialized for faces or for other familiar object categories. Nor does the LOC simply extract local three-dimensional structure from the visual scene. Rather, it appears to be more generally involved in representing complex shape.

The LOC has many of the properties expected of a visual area that subserves object recognition. The BOLD responses recorded from the LOC are larger for objects than non-objects, and the responses — at least in LOa — show both form-cue invariance and perceptual constancy for changes in size and position. The adaptation results suggest that perceptual constancy occurs locally in LOa, at a level below the resolution of current fMRI.

Current studies are limited by the spatial resolution of fMRI. It is sometimes assumed that the volume defined by fMRI — the voxel — reflects a fundamental unit of neural computation. In fact, this volume is determined solely by the spatial resolution of current fMRI scanners. Better resolution will be required before fMRI can be as useful for circuit analysis as are single-unit or multi-unit neurophysiological techniques. During natural vision, object recognition is not entirely stimulus driven, but is tightly integrated with other neural processes. Object recognition and memory, for example, do not usually operate independently. Future experiments will have to address the contributions of memory, attention and other cognitive processes to object recognition.

The studies discussed here suggest that a single region of cortex, the LOC, represents visual objects regardless of inducing cues, size or position. In humans, the local circuits mediating these effects remain unresolved at the single-neuron level. These recent studies eloquently demonstrate that, despite limited spatial resolution, fMRI is a powerful tool for exploring the neural substrates of human visual perception.

References

1. Kourtzi Z, Kanwisher N: **Cortical regions involved in perceiving object shape.** *J Neurosci* 2000, **20:**3310-3318.
2. Felleman DJ, Van Essen DC: **Distributed hierarchical processing in the primate cerebral cortex.** *Cerebral Cortex* 1991, **1:**1-47.
3. Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell, RBH: **Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging.** *Science* 1995, **268:**889-893.
4. Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady, TJ, Rosen BR, Tootell RB: **Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex.** *Proc Natl Acad Sci USA* 1995, **92:**8135-8139.
5. Kanwisher N, McDermott J, Chun MM: **The fusiform face area: a module in human extrastriate cortex specialized for face perception.** *J Neurosci* 1997, **17:**4302-4311.
6. Epstein R, Harris A, Stanley D, Kanwisher N: **The parahippocampal place area: Recognition, navigation, or encoding?** *Neuron* 1999, **23:**115-125.
7. Tanaka K: **Inferotemporal cortex and object vision.** *Annu Rev Neurosci* 1996, **19:**109-139.
8. Grill-Spector K, Kushnir T, Edelman S, Itzchak Y, Malach R: **Cue-invariant activation in object-related areas of the human occipital lobe.** *Neuron* 1998, **21:**191-202.
9. Desimone R, Albright TD, Gross CG, Bruce CJ: **Stimulus-selective properties of inferior temporal neurons in the macaque.** *J Neurosci* 1984, **4:**2051-2062.
10. Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R: **Differential processing of objects under various viewing conditions in the human lateral occipital complex.** *Neuron* 1999, **24:**187-203.
11. Miller EK, Gochin PM, Gross CG: **Habituation-like decrease in the responses of neurons in inferior temporal cortex of the macaque.** *Vis Neurosci* 1991, **7:**357-362.
12. Buckner RL, Goodman J, Burock M, Rotte M, Koutstaal W, Schacter D, Rosen B, Dale, AM: **Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI.** *Neuron* 1998, **20:**285-296.